# Demonstrating to a humanoid robot how to conduct neuropsychological tests

Duc Canh Nguyen, Frédéric Elisei, Gérard Bailly

# Demonstrating to a humanoid robot how to conduct neuropsychological tests

Duc Canh Nguyen, Gérard Bailly & Frédéric Elisei

*GIPSA-Lab, Speech & Cognition dpt., Univ. Grenoble-Alpes, France*
firstname.lastname@gipsa-lab.fr

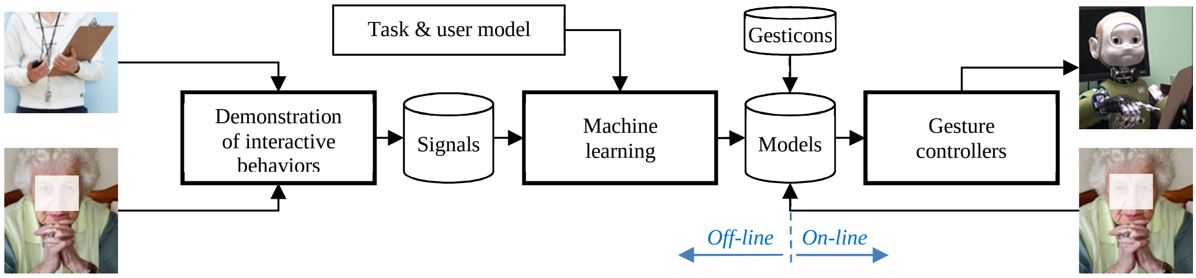


**Figure 1. The three main steps of learning interaction models by demonstration. The main contribution of this paper concerns the final part of the pipeline i.e. the gesticons and associated gesture controllers.**

## ABSTRACT

Several socially assistive robot (SAR) systems have been proposed and designed to engage people into various interactive exercises such as physical training [1], neuropsychological rehabilitation [2] or cognitive assistance [3]. While the interactive behavioral policies of most systems are scripted, we discuss here key features of a new methodology we developed in the framework of the SOMBRERO project[1] that enables professional caregivers to demonstrate a SAR how to perform the assistive tasks while giving proper instructions, demonstrations and feedbacks.

## 1 THE SOMBRERO FRAMEWORK

The three main steps of learning interaction by demonstration are given in Figure 1: we should (1) collect representative interactive behaviors from human coaches, notably when the interaction is conducted by professional coaches; (2) build comprehensive models of these overt behaviors given observed behaviors of interlocutors and a priori knowledge (task & user model, etc); and then (3) provide the target robot with appropriate gesture controllers to execute the desired behaviors.
This framework faces several problems: (1) the scaling of the human model to the interaction capabilities of the robots in terms of physical limitations (degrees of freedom) and perception, action and reasoning; (2) the drastic changes of human behaviors in front of robots or virtual agents [4]; (3) the modeling of joint interactive behaviors (4) the replay and assessment of these behaviors by the robot.
SOMBRERO proposes to solve the two first issues by enabling coaches to demonstrate human-robot interaction (HRI) via immersive teleoperation, i.e. by direct robotic embodiment. The so-called *beaming* of the gaze and lip movements of our iCub robot Nina is described in [5]. The signals in Figure 1 are thus already HRI data because the human pilot has artificially provided the SAR with cognitive skills that are adapted to the robot sensorimotor abilities.
The third issue has been addressed by Mihoub et al [6], [7]. They proposed to train statistical behavioral models that jointly map discrete multimodal events performed by the interlocutors.

## 2 THE CURRENT CONTRIBUTION

We address here the fourth issue i.e. the replay and assessment of interactive behaviors by the robot. We should in fact verify that the planned multimodal behaviors can be effectively reproduced by the target robot and that they are perceived as adequate by human interlocutors. Inadequate (e.g. speaking to somebody when looking elsewhere can be interpreted as contempt) or incomplete multimodal behaviors (e.g. looking down without moving the eyelids down can be interpreted as fear) could in fact strongly impair the communicative intents.

### 2.1 The scenario

These interviews are based on the French adaptation [8] of the Selective Reminding Test [9] named the RL/RI 16. It provides a simple and clinically useful verbal memory test for identifying loss of episodic memory in the elderly. The RL/RI 16 protocol consists in four phases: (1) the progressive learning of 16 words together with their semantic categories; (2) three successive recall tasks (free recall, complemented by an indexed-by-category recall for the unrecovered items) separated with a distractive task (reverse counting); (3) a recognition task involving the 16 items, 32 distractors (16 different words with the same semantic category and 16 true distractors) and (4) a delayed free and indexed recall (not administrated in the present study). Mnesic performance is evaluated by comparing recall rates of the subject with regards to mean & standard deviations observed within sane control population of the same age interval.

### 2.2 Interactive data

The behavioral data of the interviewer served as demonstration for the humanoid robot. Since beaming of the upper body (notably of the arms) was not available, the discrete multimodal events have been collected via semi-automatic labelling of human-human interactions (HHI). The motion of 25 retroflexive markers placed on the plexus, shoulders, head, arms, indexes and thumbs of the professional interviewer were monitored thanks to a Qualysis® system with 4 cameras. A Pertech® head-mounted monocular eyetracker also monitors the gaze of the interviewer (see Figure 2). Speech data are captured via OKMII high-quality ear microphones and are recorded synchronously with a side-view video by HD camera.

[1] See http://www.gipsa-lab.fr/projet/SOMBRERO

**Figure 2. Visual data. Left: side view from a fixed HD camera. Right: head-related view from the eyetracker scene camera. The dot superimposed to the scene camera features the current gaze fixation point.**

Each interview lasts around 20”, comprising the collection of personal records, the core RL/RI protocol and final report of performance. We analyze here a total of two hours of multimodal data for five subjects, interacting with a unique interviewer (a medical one, professionaly trained to conduct these RL/RI tests).

### 2.3 Gestural scores

Elan [10] (see Figure 3) and Praat [11] were used to semi-automatically identify speech, gaze and arm events. Behavioral models have to orchestrate these events according to the task and should be able to generate motor actions from percepts. ***Modality-specific gesture controllers have then to reproduce final motions from these discrete motor events.***

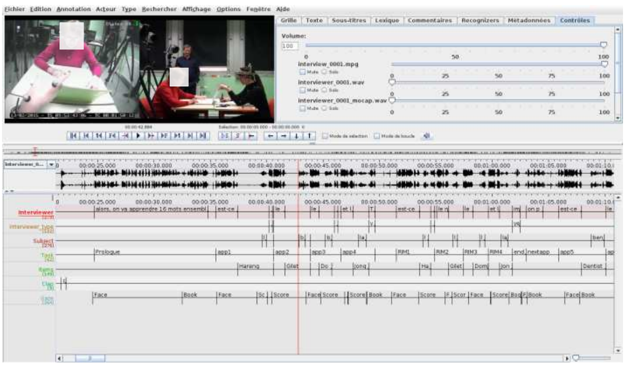

**Figure 3. Labelling gaze & speech events with Elan.**

### 2.4 Gesture controllers

***Speech.*** We transcribed speech and aligned its phonetic content with the acoustic signals uttered by both the interviewer and the subjects. The subject’s speech is mainly used to trigger scoring. The interviewer’s speech was analyzed more in depth with a special attention to prosody and in particular to backchannels [12]. The transcription together with prosodic markers is then played by the audiovisual text-to-speech synthesizer controlling Nina’s loudspeaker and facial movements [13].

***Arm gestures.*** While the human interviewer was displaying word items and scoring using sheets of paper, we decided to use tablets to display items and pretend to trigger the display and take notes (see Figure 1). Arm displacements and finger clicks of the robot are then programmed to trigger display on the subject’s tablet (show/hide items) and take notes (monitor correct responses).

***Gaze.*** We distinguish three main region of interest of the interviewer’s gaze: (1) the subject’s face; (2) the scoring tablet (i.e. scoring sheet and chronometer for original HHI); (3) the subject’s tablet (i.e. notebook for HHI). All arm gestures are performed with visuomotor supervision: since robot motion is often slower than human motion, all arm motions are preceded by one fixation towards the target if any and accompanied by gaze smooth pursuit till completion. This visuomotor supervision supersedes any other observed fixation pattern.

## 3 EVALUATION

These complex and coordinated behaviors should be perceived and interpreted correctly by subjects. We have previously shown that the morphology and appearance of effectors can strongly impair the perception of planned gestures [14]. We are thus planning to ask third parties to rate the final rendering of this multimodal score. In line with online evaluation methods deployed for audio [15] and video [16], we are planning to ask subjects to put themselves in the place of our subjects and rate the adequacy of the SAR’s behavior with regards to the subjects’ verbal behavior that they will listen to.

## 4 CONCLUSIONS

We proposed here an original framework for collecting, modelling and controlling SAR. All the building blocks are almost operational and have been evaluated separately. We plan to conduct robot-mediated HHI very soon and see what parts of this framework should be corrected. One of the key challenges is system’s adaptation. Mihoub et al [7] have shown that a subject-independent gaze model may be parametrized to adapt to specific social profiles. We will see if this approach scales to multimodal behavior planning and control.

## REFERENCES


[1] J. Fasola and M. Mataric, “A socially assistive robot exercise coach for the elderly,” *Journal of Human-Robot Interaction*, vol. 2, no. 2, pp. 3–32, 2013.

[2] J.-J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied, “Why Robots? A survey on the roles and benefits of social robots in the therapy of children with autism,” *International Journal of Social Robotics*, vol. 5, no. 4, pp. 593–618, 2013.

[3] M. E. Pollack, L. Brown, D. Colbry, C. E. McCarthy, C. Orosz, B. Peintner, S. Ramakrishnan, and I. Tsamardinos, “Autominder: An intelligent cognitive orthotic system for people with memory impairment,” *Robotics and Autonomous Systems*, vol. 44, no. 3, pp. 273–282, 2003.

[4] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, and K. L. Koay, “Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion,” *Autonomous Robots*, vol. 24, no. 2, pp. 159–178, 2008.

[5] G. Guillermo, P. Carole, F. Elisei, F. Noel, and G. Bailly, “Qualitative assesment of a beaming environment for collaborative professional activities,” in *European conference for Virtual Reality and Augmented Reality (EuroVR)*, 2015, p. 8 pages.

[6] A. Mihoub, G. Bailly, C. Wolf, and F. Elisei, “Graphical models for social behavior modeling in face-to face interaction,” *Pattern Recognition Letters*, vol. 74, pp. 82–89, 2016.

[7] A. Mihoub, G. Bailly, and C. Wolf, “Learning multimodal behavioral models for face-to-face social interaction,” *Journal on Multimodal User Interfaces (JMUI)*, vol. 9, no. 3, pp. 195–210, 2015.

[8] M. Van der Linden, F. Coyette, J. Poitrenaud, M. Kalafat, F. Calicis, C. Wyns, and S. Adam, “L’épreuve de rappel libre / rappel indicé à 16 items (RL/RI-16),” in *L’évaluation des troubles de la mémoire : présentation de quatre tests de mémoire épisodique avec leur étalonnage*, M. Van der Linden, Ed. Marseille, France: Solal, 2004, pp. 25–47.

[9] E. Grober and H. Buschke, “Genuine memory deficits in dementia,” *Developmental Neuropsychology*, vol. 3, pp. 13–36, 1987.

[10] B. Hellwig and D. Uytvanck, “EUDICO Linguistic Annotator (ELAN) Version 2.0.2 manual,” Max Planck Institute for Psycholinguistics, Report, 2004.

[11] P. Boersma and D. Weenink, *Praat, a System for doing Phonetics by Computer, version 3.4*. 1996.
[12] Bailly, Gérard, Elisei, Frédéric, Juphard, Alexandra, and Moreau, Olivier, “Quantitative analysis of backchannels uttered by an interviewer during neuropsychological tests,” in *Interspeech*, San Francisco, CA, submitted.
[13] A. Parmiggiani, M. Maggiali, L. Natale, F. Nori, A. Schmitz, N. Tsagarakis, J. S. Victor, F. Becchi, G. Sandini, and G. Metta, “The design of the iCub humanoid robot,” *International Journal of Humanoid Robotics*, vol. 9, no. 4, p. 24 pages, 2012.
[14] F. Foerster, G. Bailly, and F. Elisei, “Impact of iris size and eyelids coupling on the estimation of the gaze direction of a robotic talking head by human viewers,” in *Humanoids*, Seoul, Korea, 2015.
[15] M. Hansen and B. Kollmeier, “Continuous assessment of time-varying speech quality,” *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2888–2899, 1999.
[16] Hamberg, Roelof and de Ridder, Huib, “Continuous assessment of perceptual image quality,” *Journal of the Optical Society of America*, vol. 12, no. 12, pp. 2573–2577, 1995.